Identification of disease-promoting stromal components by comparative proteomic and transcriptomic profiling of canine mammary tumors using laser-capture microdissected FFPE tissue

NEOPLASIA
www.neoplasia.com

Volume 23 Number xxx Month 2021 pp. 400–412

Identification of disease-promoting stromal components by comparative proteomic and transcriptomic profiling of canine mammary tumors using laser-capture microdissected FFPE tissue ☆, ☆☆

Amiskwia Pöschel [a]; Erin Beebe [a]; Laura Kunz [b]; Parisa Amini [a]; Franco Guscetti [c]; Alexandra Malbon [d]; Enni Markkanen [a,*]

[a] Institute of Veterinary Pharmacology and Toxicology, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland
[b] Functional Genomics Center Zürich, ETH Zürich/University of Zurich, Zurich, Switzerland
[c] Institute of Veterinary Pathology Vetsuisse Faculty, University of Zurich, Zürich, Switzerland
[d] The Royal (Dick) School of Veterinary Studies and The Roslin Institute Easter Bush Campus, Midlothian, Scotland

Abstract

Cancer-associated stroma (CAS) profoundly influences progression of tumors including mammary carcinoma (mCA). Canine simple mCA represent relevant models of human mCA, notably also with respect to CAS. While transcriptomic changes in CAS of mCA are well described, it remains unclear to what extent these translate to the protein level. Therefore, we sought to gain insight into the proteomic changes in CAS and compare them with transcriptomic changes in the same tissue. To this end, we analyzed CAS and matched normal stroma using laser-capture microdissection (LCM) and LC-MS/MS in a cohort of 14 formalin-fixed paraffin embedded (FFPE) canine mCAs that we had previously characterized using LCM-RNAseq. Our results reveal clear differences in protein abundance between CAS and normal stroma, which are characterized by changes in the extracellular matrix, the cytoskeleton, and cytokines such as TNF. The proteomics- and RNAseq-based analyses of LCM-FFPE show a substantial degree of correlation, especially for the most deregulated targets and a comparable activation of pathways. Finally, we validate transcriptomic upregulation of LTBP2, IGFBP2, COL6A5, POSTN, FN1, COL4A1, COL12A1, PLOD2, COL4A2, and IGFBP7 in CAS on the protein level and demonstrate their adverse prognostic value for human breast cancer. Given the relevance of canine mCA as a model for the human disease, our analysis substantiates these targets as disease-promoting stromal components with implications for breast cancer in both species.

Neoplasia (2021) 23, 400–412

Keywords: Tumor microenvironment, Cancer-associated fibroblasts, Comparative oncology, Breast cancer, Tumor stroma, Canine mammary carcinoma

Introduction

The microenvironment that surrounds cancer cells, the so-called cancer-associated stroma (CAS), has profound effects on the development and survival of tumor cells, thereby strongly influencing clinical aspects of the disease [1–3]. Indeed, CAS has been shown to modulate most of the hallmarks of cancer in a wide variety of tumors, including breast cancer [3]. CAS is composed of a wide array of different non-malignant cells (among them fibroblasts, immune cells, vascular cells, adipocytes, and others) that are embedded into an insoluble extracellular matrix (ECM). Reprogramming of normal stroma to CAS is strongly driven by the adjacent cancer cells. By producing a range of growth factors and proteases which modify the surrounding stromal environment (reviewed in [4]), cancer cells remodel the cellular surroundings to their own advantage. To date however, our understanding regarding CAS reprogramming in patient samples and the molecular dialogue between CAS and cancer cells remains incomplete.

Based on clinical, histological and molecular similarities, canine simple mammary carcinomas (mCA) are considered an excellent model for human breast cancer, also because they overcome several of the limitations of xenograft or genetically modified rodent tumor models [5–7]. Indeed, canine simple mCA not only emulate the biology of human mCA but also feature many of the genomic aberrations found therein [5,6]. mCA are the most

* Corresponding author.
E-mail address: enni.markkanen@vetpharm.uzh.ch (E. Markkanen).
☆ Funding: This study was financially supported by the Promedica Stiftung Chur to EM.
☆☆ Conflicts of interest: None.
Received 12 January 2021; received in revised form 1 March 2021; accepted 2 March 2021

https://doi.org/10.1016/j.neo.2021.03.001

frequent tumors in both women and intact female dogs [8]. Canine simple mCA are malignant epithelial neoplasms that infiltrate the surrounding tissue, whereby they induce a strong stromal response, and can also give rise to metastases [9]. Importantly, the similarities between human and canine mCA are not only limited to the tumor cells, but also extend to reprogramming of CAS. By analyzing CAS and normal stroma from formalin-fixed paraffin embedded (FFPE) breast cancer tissue using laser-capture-microdissection (LCM) through RNA sequencing (RNAseq) and quantitative PCR (RT-qPCR), we have recently demonstrated the existence of strong molecular homologies in stromal reprogramming between human and canine mCA [10–14].

While transcriptomic changes in CAS of both human and canine mCA are beginning to be understood, it remains unclear to what extent these differences in mRNA abundance actually translate to the protein level. Thus far, analytic approaches of CAS reprogramming in both human and canine patient samples of mCA have been mostly focused on analysis of RNA, therefore reflecting the transcriptional state of the tissue [11,12,15,16]. However, because RNA levels and protein abundance do not necessarily correlate [17–19], it remains entirely unclear to what extent the observed transcriptional changes translate to the protein level in both human and canine mCA. To the best of our knowledge, there is only one report describing proteomic changes during stromal reprogramming in fresh-frozen samples of human mCA [20] and none for canine mCA. This striking shortage of data on a highly relevant aspect of tumor biology warrants further investigation.

FFPE tissue represents a huge resource of patient material that is routinely prepared in pathology departments and can be easily stored for decades. However, FFPE negatively impacts on the quality and quantity of macromolecules that can be isolated from these tissues, making analysis of RNA or protein from such tissue challenging. Having already established RNA analysis of microdissected areas from such tissue, we set out to define proteomic changes during stromal reprogramming in archival FFPE patient samples and to compare them with transcriptomic changes in the same tissue. To this end, we isolated CAS and matched normal stroma using LCM from a cohort of 14 archival FFPE canine mCAs that we had previously characterized using LCM-RNAseq [11], and analyzed them by LC-MS/MS.

Material and methods

Selection of cases for LCM

CAS and matched normal stroma were isolated from fourteen dog mammary carcinoma samples that were provided by the Institute of Veterinary Pathology of the Vetsuisse Faculty Zürich. All samples were formalin-fixed, paraffin-embedded tissue samples either from the Small Animal Hospital of Zurich or external cases sent in by veterinarians practicing in Switzerland. Case selection had been performed as part of an earlier research project that analyzed the transcriptomic changes in cancer-associated stroma of 15 cases in detail using RNAseq [11]. The identical samples were used for this project to be able to compare the proteomic changes with the results obtained by RNAseq. Original case number 7 had to be excluded as the paraffin block was not available anymore. Details regarding selection criteria are described in [13]. Criteria for case selection included female dogs, simple mammary adenoma, and sufficient tumor stroma content for tissue isolation. Paraffin blocks were routinely kept at room temperature. Tissue processing for LCM was performed as described in [10]. All cases were reviewed by a veterinary pathologist (Alexandra Malbon, Franco Guscetti). Table 1 provides clinical details, such as age and breed of each patient, sample age and tumor type, for all cases included in the study.

Laser-capture microdissection (LCM)

Laser-capture microdissection was performed using the ArcturusXT Laser Capture Microdissection System (Thermo Scientific) and the Arcturus CapSure Macro LCM Caps (Life Technologies). Areas of normal or cancer-associated stroma from the specimen were identified and isolated according to the manufacturer's protocol and the criteria described in [11,13]. Isolation of areas of interest was verified by microscopic examination of the LCM cap as well as the excised region after microdissection. We collected 2 caps of CAS and normal stroma each. After excision, the filled caps containing tissue were put on a 1.5 ml centrifuge tube (EppendorfSafe-Lock tubes) and frozen at −20°C, until further processing.

Sample preparation for proteomic analysis

For protein extraction, sterile blades and forceps were used to peel off the two thermoplastic membranes containing the captured cells from the cap, which were then transferred into a sterile EppendorfSafe-Lock tube. The excised tissue pieces were rehydrated by adding 900 μl of heptane and incubating for 10 min at 30°C in a thermomixer (800 rpm). After centrifugation (20 000 × g, 10 min) the heptane was removed, and this step was repeated. Subsequently the membranes were washed with 900 μl of 100% ethanol (5 min, RT, 1000 rpm), 200 μl of 90% ethanol (5 min, RT, 1000 rpm) and 200 μl of 75% ethanol (5 min, RT, 1000 rpm), and then stored at −80°C overnight. Further sample preparation was performed by using a commercial iST Kit (PreOmics, Germany) with an updated version of the protocol. Briefly, the tissues were solubilized in 'Lyse' buffer, boiled at 95°C for 60 minutes and processed with High Intensity Focused Ultrasound (HIFU) for 2 times 60 seconds setting the ultrasonic amplitude to 85%. After 1:1 dilution with dH2O, the protein concentration was estimated using the Qubit Protein Assay Kit (Life Technologies, Zurich, Switzerland). Samples were then transferred to the cartridge and digested by adding 50 µl of the 'Digest' solution. After 3 hours of incubation at 37°C the digestion was stopped with 100 µl of Stop solution. The solutions in the cartridge were removed by centrifugation at 3800 × g, while the peptides were retained by the iST-filter. Finally, the peptides were washed, eluted, dried and re-solubilized in 20 µL of MS-solution (3% acetonitrile, 0.1% formic acid) for LC-MS-Analysis.

Liquid chromatography-mass spectrometry analysis

Mass spectrometry analysis was performed on a Q Exactive HF-X mass spectrometer (Thermo Scientific) equipped with a Digital PicoView source (New Objective) and coupled to a M-Class UPLC (Waters). Solvent composition at the two channels was 0.1% formic acid for channel A and 0.1% formic acid, 99.9% acetonitrile for channel B. For each sample 2 μl of peptides were loaded on a commercial MZ Symmetry C18 Trap Column (100 Å, 5 µm, 180 µm × 20 mm, Waters) followed by nanoEase MZ C18 HSS T3 Column (100 Å, 1.8 µm, 75 µm × 250 mm, Waters). The peptides were eluted at a flow rate of 300 nL/min by a gradient from 8 to 27% B in 85 min, 35% B in 5 min and 80% B in 1 min. Samples were acquired in a randomized order. The mass spectrometer was operated in data-dependent mode (DDA), acquiring a full-scan MS spectra (350−1400 m/z) at a resolution of 120 000 at 200 m/z after accumulation to a target value of 3 000 000, followed by HCD (higher-energy collision dissociation) fragmentation on the twenty most intense signals per cycle. HCD spectra were acquired at a resolution of 15 000 using a normalized collision energy of 25 and a maximum injection time of 22 ms. The automatic gain control (AGC) was set to 100 000 ions. Charge state screening was enabled. Singly, unassigned, and charge states higher than seven were rejected. Only precursors with intensity above 250 000 were selected for MS/MS. Precursor masses previously selected for MS/MS measurement were

Table 1

Clinical data from canine patients with simple mammary carcinoma included in this study.

Case #	Gender	Breed	Age [years]	Simple carcinoma	Sub type of simple carcinoma
1	f	Basset	12	yes	tubular
2	f	Vizsla	10	yes	cystic-papillary
3	f	Samoyed	5	yes	tubulopapillary
4	f	Maltese	14	yes	tubular
5	f	Tibetan Terrier	12	yes	tubular
6	f/n	West Highland White Terrier	12	yes	tubular-solid
8	f	Chihuahua	8	yes	tubulopapillary
9	f/n	Bracke	9	yes	cribriform
10	f/n	n.d.	13	yes	tubular
11	f/n	Appenzell Mountain Dog	6	yes	tubular
12	f	Boxer	9	yes	tubulopapillary
13	f	n.d.	4	yes	cystic-papillary
14	f/n	Beagle	13	yes	tubular
15	f/n	Chihuahua	10	yes	tubular

f, female; f/n, female neutered, n.d., not disclosed; age, age of patient at excision of tumor.

excluded from further selection for 30 s, and the exclusion window was set at 10 ppm. The samples were acquired using internal lock mass calibration on m/z 371.1012 and 445.1200. The mass spectrometry proteomics data were handled using the local laboratory information management system (LIMS) [21].

Protein identification and label free protein quantification

The acquired raw MS data were processed by MaxQuant (version 1.6.2.3), followed by protein identification using the integrated Andromeda search engine [22]. Spectra were searched against a Uniprot Canis lupus familiaris reference proteome (taxonomy 9615, canonical version from 2019-07-29), concatenated to its reversed decoyed fasta database and common protein contaminants. Carbamidomethylation of cysteine was set as fixed modification, while methionine oxidation and N-terminal protein acetylation were set as variable. Enzyme specificity was set to trypsin/P allowing a minimal peptide length of 7 amino acids and a maximum of two missed cleavages. MaxQuant Orbitrap default search settings were used. The maximum false discovery rate (FDR) was set to 0.01 for peptides and 0.05 for proteins. Label free quantification was enabled and a 2 minute window for match between runs was applied. In the MaxQuant experimental design template, each file is kept separate in the experimental design to obtain individual quantitative values.

Bioinformatics data analysis

Protein fold changes were computed based on peptide intensity values reported in the MaxQuant generated peptides.txt file, using linear mixed-effects models [23]. Pre-processing of the peptide intensities reported in the peptides.txt file was performed as follows: intensities equal zero were removed, non-zero intensities were log2 transformed and modified using robust z-score transformation. Afterwards, for each protein, a mixed-effects model was fitted to the peptide intensities, and fold changes and p-values were computed based on this model for each contrast [24]. For a contrast, all the protein p-values are adjusted using the Benjamini and Hochberg procedure to obtain the false discovery rate (FDR) [25]. To estimate fold-changes for proteins for which mixed-effects model could not be fit because of an excess in missing measurements, the following procedure was applied: First, the mean intensity for all peptides for each condition was computed. For the proteins with no measurements, we imputed the peptide intensities using the mean of the 10% smallest average peptide intensities computed in step one. Afterwards, contrasts (differences between conditions) were computed for each peptide and finally, the median of the peptide estimates was used to provide a per protein fold change estimate.

Data analysis

All statistical analysis was performed using the ANOVA test. For the gene set enrichment analysis and KEGG pathway analysis, the tool WebGestalt (http://www.webgestalt.org) was used. For the comparison of the transcriptomic and proteomic data set, the online tool Shiny App (http://fgcz-shiny.uzh.ch/fgcz_multiOmicsAnalysis_app/) run by the Functional Genomics Center Zurich, was used. Dog genes (CanFam3.1) were converted to human orthologues using Ensembl bioMart (release 100) prior to analysis with MetaCore [72]. For the pathway analysis, the web tool MetaCore from Clarivate Analytics was used (https://portal.genego.com).

Survival analysis in human patient cohorts

Correlation between protein expression and survival of human breast cancer patients was assessed using the online tool "KM plotter for breast protein" https://kmplot.com/analysis/ using the standard settings for overall survival (auto select best cutoff). If present in more than one dataset, only data from targets that showed the same expression trend in all datasets analyzed were included.

Immunofluorescence and immunohistochemistry

FFPE tissue sections (2 μm thickness) were mounted on positively charged SuperFrost Plus slides (Thermo Fisher Scientific) and dried overnight at 37°C. Deparaffinization of the slides was performed with four xylene baths for 5 min each using the Tissue-Tek Prisma and Film (Sysmex), and slides were rehydrated using a degressive alcohol series of 100% ethanol, 95% ethanol, 70% ethanol and distilled water. All immunofluorescence sections were counterstained with 0.1 mg/ml DAPI (Thermo Fisher Scientific) at room temperature for 15 min to visualize nuclei. The anti-CollagenIV-alpha1 antibody was from BioRad (2150-1470), and was used at 1:50 dilution at 4°C overnight after basic antigen retrieval by 20 min incubation in pH 9.0 EDTA buffer in a pressure cooker set to 98°C. The secondary antibody used was goat anti mouse (Invitrogen A11005, 594), diluted 1:400, and incubated for 1 hour at room temperature. The anti-fibronectin antibody was from Abcam (ab2413), the reaction was run in a Ventana Autostainer using the

CC2 st. antigen retrieval option, the primary antibody was diluted 1:200 and incubated for 1 h at RT, the reaction was visualized with the RedMap Kit.

Graphical display of results

For visual representation and statistical analysis, the programs GraphPad Prism (www.graphpad.com), Shiny App and MetaCore were used.

Results

Proteomic profiling of matched CAS and normal stroma in canine simple mammary carcinoma isolated using laser-capture microdissection of FFPE samples

To perform proteomic profiling of CAS reprogramming in canine simple mCA, we concurrently isolated CAS and matched 'normal' stroma (i.e. stroma that was found adjacent to unaltered mammary glands) using LCM from 14 clinical FFPE specimens in which we had previously analyzed CAS reprogramming using RNAseq [11]. Representative images for tissue isolation and patient characteristics for all cases included in the study can be found in Table 1 and Supplementary Fig. 1. LC-MS/MS analysis identified 10'858 different peptides corresponding to 2'371 different proteins in the 14 pairs of normal stroma and CAS that were analyzed (Fig. 1A–C, Supplementary Table 1). When considering only targets for which 2 or more peptides were detected, 1'591 different proteins could be identified (Fig. 1D and E). Overall, samples of CAS tended to yield a higher number of identified proteins (Fig. 1B, C and E).

Principal component analysis (PCA) using all identified proteins revealed a clear separation of CAS from normal stroma, suggesting the difference between normal stroma and CAS to be among the main sources of variability in the data (Fig. 2A). Differential expression analysis of the proteins with an adjusted p-value cut-off of 0.05 and a fold-change threshold of 2 revealed 282 proteins to be significantly differentially expressed between CAS and normal stroma. Of these, 240 were up- and 42 were down-regulated in CAS compared to normal stroma (Fig. 2B, Table 2, and Supplementary Table 1). Over-representation analysis of Gene Ontology (GO) terms associated with biological processes, cellular components, and molecular functions revealed changes in the following categories: muscle system and contraction, tumor necrosis factor (TNF) production, organization of the extracellular matrix, contraction and organization of actin filaments, extracellular space, and integrin binding (Fig. 2C). Overall, these results demonstrate the feasibility of our approach to identify proteins from microdissected FFPE sections and reveal clear differences in protein abundance between CAS and normal stroma from canine mCA. The observed changes are characterized by alterations in proteins associated with the extracellular matrix, the cytoskeleton, and cytokines such as TNF. As such, this is the first dataset to assess proteomic changes between CAS and normal stroma of canine mammary carcinomas.

Comparison of proteomic and transcriptomic profiling of stromal reprogramming reveals comparable results

To understand how the observed proteomic changes in CAS reprogramming relate to the underlying transcriptomic changes in the tissue, we compared the proteomic dataset to RNAseq data that was previously obtained using the same patient samples and identical approach to isolate CAS and normal stroma [11]. To compare the two datasets, we identified the Entrez Gene ID for the proteins for which at least two peptides had been detected and juxtaposed these to the Gene IDs of the transcriptomic dataset. 1'377 proteins could be annotated with a Gene ID, compared to 18'494 targets that were detected using RNAseq. The majority of the proteomic dataset overlapped with the transcriptomic one, with 1'367 targets that were identified by both approaches (Fig. 3A). 10 proteins were not identified using transcriptomics, most likely due to annotation deficiencies between different databases. RNAseq identified 17'132 transcripts that were not detected at the protein level (Fig. 3A). A list of the 10 targets uniquely identified by proteomics can be found in Supplementary Table 2. Comparative studies have found correlations between mRNA and protein levels in model organisms to range from relatively weak and uncertain to moderately positive [17,18,26]. In accordance with this, correlation analysis of the expression changes between CAS and normal stroma for the targets identified through both transcriptomics and proteomics for all data points (adjusted p-value threshold 1) revealed a correlation coefficient of R=0.3626 between transcriptomics and proteomics (Fig. 3B). Restricting the analysis to the significantly changing targets by setting the threshold of the adjusted p-value to 0.05 increased the correlation coefficient to R=0.5451 (Fig. 3C). Furthermore, linear regression analysis of the top 20 up- and down-regulated targets from the RNAseq dataset for which the respective proteomic data was available revealed a significant positive correlation with R=0.7754 (Fig. 3D and F). A similar picture emerged upon comparison of the top 20 up- and down-regulated protein targets with their RNAseq counterparts (Fig. 3E and G).

To understand which pathways were enriched in both datasets, we performed MetaCore analysis (Clarivate Analytics) of the human orthologues obtained using Ensembl bioMart application of all proteomic and transcriptomic targets. The top 10 pathways and the proteins associated with them can be found in Fig. 4A and B. TGF-beta related signaling was evident in 4/10 pathways (no. 1, 2, 5 and 6), suggesting a strong involvement of TGF-beta signaling in stromal reprogramming in canine mCA. Pathway number 3 identified IL1 beta and endothelin-1 induced fibroblast migration and extracellular matrix production, and HIF-1 signaling was found as number 4. Furthermore, WNT/beta-catenin signaling was pathway number 7, followed by regulation of epithelial-to-mesenchymal transition (EMT), the involvement of the IGF family in invasion and metastasis and glucocorticoid-related pathways in positions 8, 9 and 10. In summary, these results suggest that, while the proteomics-based analysis of LCM-FFPE tissue yields a smaller set of detected targets, the protein and transcript levels show a substantial degree of correlation, especially for the most deregulated targets. Moreover, the results from both approaches reveal a comparable picture with regards to pathway activation during stromal reprogramming in canine mCA.

Proteomic CAS dataset validates deregulation of stromal targets with prognostic value for human breast cancer

CAS can strongly influence both development and progression of tumors. While canine simple mCA are malignant tumors that can give rise to metastases, canine simple mammary adenomas are the benign non-infiltrative counterpart to these neoplasms. It currently remains unclear whether such benign adenomas can progress to become malignant mCA. To identify genes that potentially influence malignant development of a tumor, we have previously compared transcriptomic changes in CAS from malignant canine mCAs to adenoma-associated stroma (AAS) from benign mammary adenomas [12]. This identified a list of interesting potential disease-promoting stromal targets that are strongly upregulated in CAS from malignant mCA compared to AAS from benign adenomas, or normal stroma. However, it remains unclear whether these transcriptional changes actually translate to altered protein levels. To understand whether these potential disease-promoting stromal targets are upregulated at the protein level in CAS, we analyzed the expression of the top 21 targets of this list (Supplementary Table 3) for which proteomic data was obtained. Indeed, in accordance with the transcriptomic upregulation, all of these 21 targets were also significantly upregulated on protein level in CAS versus normal stroma (Fig. 5A, Supplementary Table 3). We next aimed to investigate the

A

Number of peptides and proteins identified in all samples	
No. of Peptide IDs	No. of Protein IDs
10'858	2'371

D

Number of proteins identified in all samples with ≥ 2 peptides
No. of Protein IDs
1'591

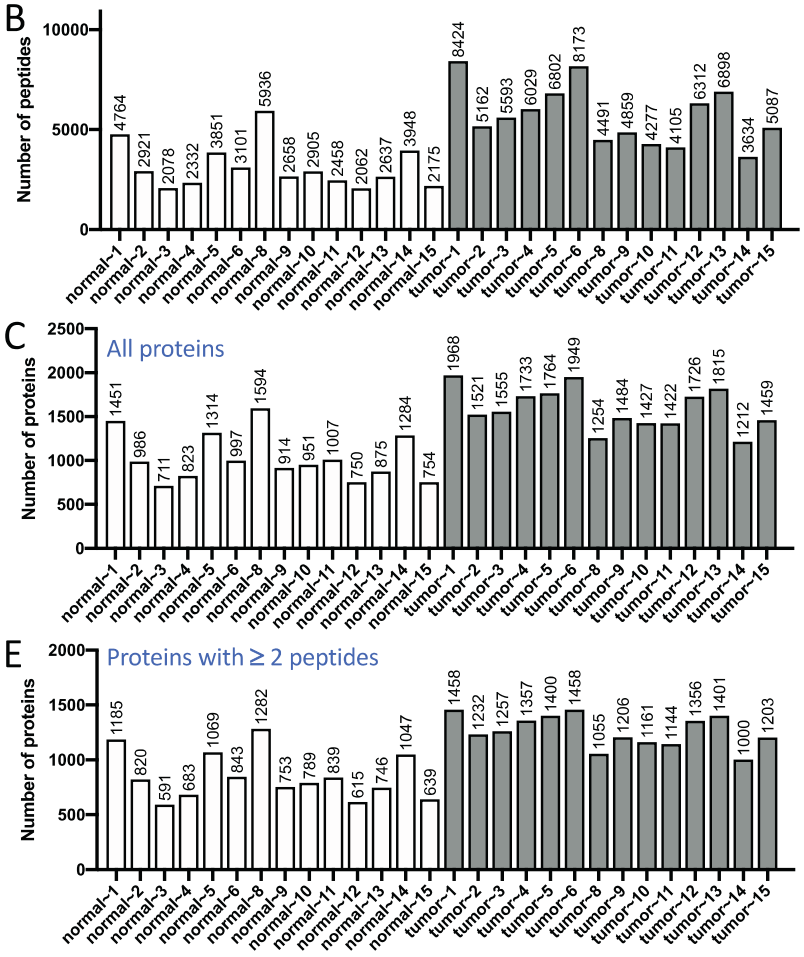


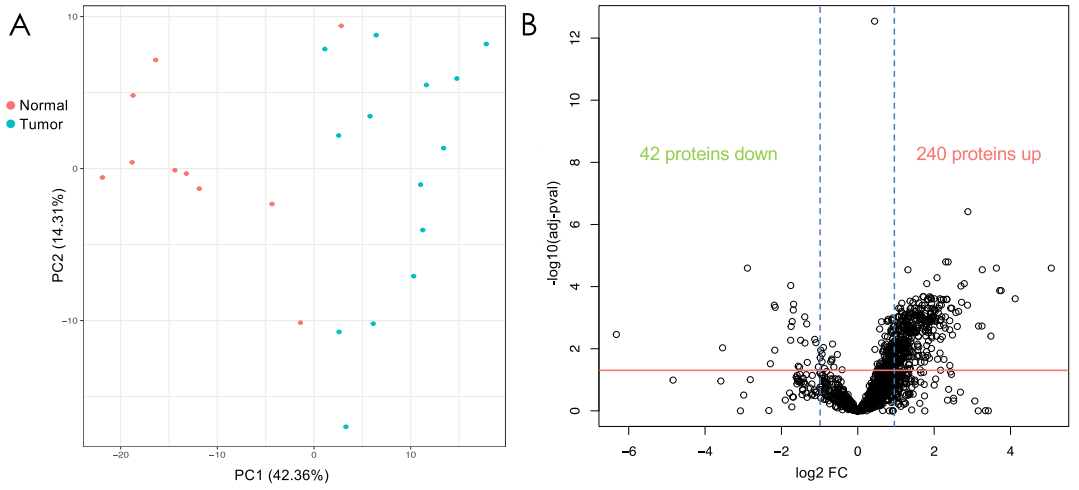
Fig. 1. Overview of number of peptides and proteins identified in CAS and normal stroma from 14 canine mammary carcinomas. A–C) Overview of all detected peptides and proteins in the 14 paired samples. A) Summary of findings. B and C) Total number of peptides (B) and proteins (C) identified in the individual samples. D and E) Overview of identified proteins for which at least 2 peptides were detected. D) Summary of findings. E) Total number of proteins identified in the individual samples. The Y-axis indicates the number of proteins per sample, the X-axis indicates the sample. Normal = normal stroma, tumor = CAS.

association of the expression of these 21 proteins with survival. Given the relevance of CAS from canine mCA for human breast cancer [11] and the lack of long-term survival data for the canine cases that were included in this study, we utilized the Kaplan-Meier Plotter (kmplot.com/analysis/) for breast protein. This online-tool allows to assess the effect of protein levels in bulk tumor samples on survival in breast cancer patients [27], based on previously published datasets [28–30]. For three of these proteins (SFRP2, SLPI, and SORCS2), there was no expression data available. Strikingly, we found high expression to be strongly correlated with worse overall survival for 10 out of the remaining 18 proteins (LTBP2, IGFBP2, COL6A5, POSTN, FN1, COL4A1, COL12A1, PLOD2, COL4A2, and IGFBP7; Fig. 5B–K)). Additionally, the five proteins BGN, CDH11, COL15A1, COL8A1, and COL8A2 showed the same trend, but did not reach the significance threshold of $p = 0.05$ (Supplementary Fig. 2). Finally, we validated the increased expression for Collagen IV by immunofluorescence and for Fibronectin by immunohistochemistry in CAS compared to normal stroma on the protein level (Fig. 5L and M).

Hence, by validating the deregulation of previously identified differentially expressed transcripts on protein level, our proteomic analysis of CAS reprogramming in canine mCA further substantiates these targets as disease-modulating stromal components with implications for breast cancer in both humans and dogs.

Discussion

CAS is well accepted to play a central role in initiation and progression of cancer [1–3]. While CAS reprogramming has been analyzed on the transcriptional level, it remains unclear to what extent transcriptomic changes therein translate to the protein level, and what the proteomic landscape of stromal reprogramming looks like in cancer patient samples. Here, we harness



C

GO: Biological processes

Cluster 1: System Processes			
Term	Gene Set ID	Normalized Enrichment Score	FDR
Muscle system process	GO:0003012	-2.3492	0.0024524
Muscle contraction	GO:0006936	-2.3268	0.0018393

Cluster 2: Cytokine Production and Regulation			
Term	Gene Set ID	Normalized Enrichment Score	FDR
Tumor necrosis factor superfamily cytokine production	GO:0071706	1.9892	0.010814
Tumor necrosis factor production	GO:0032640	1.9892	0.010814
Regulation of tumor necrosis factor superfamily cytokine production	GO:1903555	1.9892	0.010814
Regulation of tumor necrosis factor production	GO:0032680	1.9892	0.010814

Cluster 3: Extracellular Structure Organization			
Term	Gene Set ID	Normalized Enrichment Score	FDR
Extracellular matrix organization	GO:0030198	1.9188	0.022060

GO: Cellular components

Cluster 4: Cytoskeleton			
Term	Gene Set ID	Normalized Enrichment Score	FDR
Actin filament	GO:0005884	1.8541	0.029294
Actin filament bundle	GO:0032432	1.8286	0.016926
Contractile actin filament bundle	GO:0097517	1.7757	0.032549
Stress fiber	GO:0097517	1.7757	0.032549
Actomyosin	GO:0042641	1.7690	0.032549

Cluster 5: Cellular Anatomical Entity			
Term	Gene Set ID	Normalized Enrichment Score	FDR
Extracellular region	GO:0005576	1.8714	0.039059
Extracellular region part	GO:0044421	1.8362	0.024593
Extracellular space	GO:0005615	1.8319	0.018445

GO: Molecular functions

Cluster 6: Binding			
Term	Gene Set ID	Normalized Enrichment Score	FDR
Integrin binding	GO:0005178	1.9272	0.0084169

Fig. 2. Proteomic analysis of CAS and normal stroma from canine simple mCA reveals strong stromal reprogramming. A) PCA plot of 14 pairs of CAS and normal stroma. PCA was performed using all identified proteins. Normal = normal stroma, tumor = CAS. B) Volcano plot highlighting proteins that are differentially expressed between CAS and normal stroma. Red line=adjusted p-value 0.05, X-axis = estimate = log2(FC). Vertical blue dashed lines indicates cutoff-fold change of 2. P-values were calculated using the ANOVA test; C) Differently regulated biological processes, cellular components, and molecular functions between CAS and normal stroma identified through gene set enrichment analysis. Term = affected gene set; ID = gene ontology accession number; normalized enrichment score = log2 (FC) in CAS compared to normal stroma; FDR = false discovery rate, indicating the significance of overlap between gene set enrichment in CAS.

the validity of spontaneous canine simple mCA as a relevant model of human mCA, also with respect to CAS reprogramming, to gain insight into the proteomic landscape of stromal reprogramming in both human and canine breast cancer. By analyzing microdissected FFPE tissues using LC-MS/MS, we present here the first dataset to assess proteomic changes between CAS and normal stroma of canine mCA. Overall, our results demonstrate the feasibility to identify proteins from microdissected FFPE sections of patient samples and reveal clear differences in protein abundance between CAS and normal stroma in canine mCA. These results will contribute to a better understanding of the involvement of stromal genes in development and progression of canine mCA, serve as a basis for further mechanistic studies, and have the potential to reveal novel prognostic markers as well as therapeutic targets.

Archival FFPE patient samples represent a huge resource of patient material that is routinely prepared in pathology departments and can be easily stored for decades. However, FFPE negatively impacts on the quality and quantity of macromolecules that can be isolated from these tissues, making analysis of RNA or protein from such tissue challenging. Despite these difficulties, recent years have brought a big improvement in extraction and analysis methods, rendering the analysis of RNA or proteins from FFPE tissue possible while also demonstrating a good consistency between results from FFPE and fresh-frozen tissue [31–33]. Bulk analysis of fresh-frozen human breast cancer has revealed significant differences between malignant and matched non-tumor tissue [34]. The positive aspects in terms of quality of extracted macromolecules when using fresh-frozen tissue are

Table 2

Top 25 up- and down-regulated proteins in CAS compared to normal stroma.

Uniprot ID	Gene	Protein name	Adjusted *P* value	Log^2 (FC)
J9NY38	COL8A2	Collagen type VIII alpha 2 chain	1.22E-06	4.91
E2RS46	FBN2	Fibrillin 2	2.94E-02	4.38
J9P6G0	COL12A1	Collagen type XII alpha 1 chain	1.14E-288	3.96
F1Q2Y5	ZG16B	Zymogen granule protein 16B	-	3.83
F1PBI6	THBS1	Thrombospondin-1	1.05E-64	3.54
F1P6H7	FN1	Fibronectin	1.40E-10	3.45
E2QYZ4	LOXL3	Lysyl oxidase homolog	1.86E-03	3.39
F1PQD5	COL11A2	Collagen type XI alpha 2 chain	-	3.36
E2RGR7	CCDC80	Coiled-coil domain containing 80	1.52E-02	3.32
E2R1E3	PAPSS2	3′-phosphoadenosine-5′-phosphosulfate synthase	5.06E-04	3.26
F1P916	TPM1	Tropomyosin 1	1.66E-02	3.26
J9P5F1	USO1	General vesicular transport factor p115	6.78E-03	3.19
F6V544	PDLIM4	PDZ and LIM domain 4	4.91E-05	3.17
G1K2D8	BGN	Biglycan	6.54E-68	3.04
E2QUL2	VAMP8	Vesicle-associated membrane protein 8	-	2.90
F1Q2F7	COL6A5	Collagen type VI alpha 5 chain	2.25E-18	2.90
F1PEN8	BCHE	Carboxylic ester hydrolase	-	2.89
E2QSK4	CPXM1	Carboxypeptidase X, M14 family member 1	6.05E-09	2.83
F1PLN5	FMOD	Fibromodulin	1.93E-32	2.71
F6UTF3	THBS2	Thrombospondin 2	2.69E-09	2.67
E2R7R1	ISG15	ISG15 ubiquitin like modifier	1.56E-02	2.66
F6XLB5	PDLIM7	PDZ domain-containing protein	5.49E-08	2.66
D3YJ60	CHI3L1	Chitinase-3-like protein 1	3.73E-06	2.65
E2RBA8	PDLIM5	PDZ and LIM domain 5	3.00E-12	2.63
F1PEQ5	IGFBP5	Insulin-like growth factor-binding protein 5	5.62E-11	2.57
J9PAT9	HAPLN1	Hyaluronan and proteoglycan link protein 1	1.80E-07	-6.39
J9NVT5	-		1.79E-03	-6.39
F1PV96	MYL1	Myosin light chain 1	6.24E-01	-4.56
F6Y2H4	SERPINE2	Serpin family E member 2	1.05E-03	-3.36
F1PPC2	LGALS7	Galectin	4.75E-09	-2.77
E2RHU8	CALML5	Calmodulin-like 5	-	-2.43
J9NYC0	MFAP4	Microfibril-associated protein 4	5.10E-06	-2.33
E2RPQ6	DPT	Uncharacterized protein	1.14E-10	-2.24
E2RNR0	OGN	Mimecan	2.58E-46	-2.20
F1PGU1	MYLPF	Myosin light chain, phosphorylatable, fast skeletal muscle	-	-2.05
J9NV52	TNNT1	Troponin T, slow skeletal muscle	-	-2.01
Q29393	DCN	Decorin	1.74E-48	-1.95
E2RPS1	V4G	C-type lectin domain family 4 member G	3.97E-04	-1.91
E2R837	UBA2	SUMO-activating enzyme subunit 2	3.61E-02	-1.85
F1PKX3	F13A1	Coagulation factor XIII A chain	2.56E-15	-1.71
F1Q3I5	COL1A1	Collagen alpha-1(I) chain	2.31E-45	-1.68
F1PS24	COL2A1	Collagen type II alpha 1 chain	2.17E-02	-1.67
J9P0L0	COL3A1	Collagen type III alpha 1 chain	6.46E-09	-1.67
P15944	N/A	Tryptase	1.39E-02	-1.60
J9NX65	SBSN	Uncharacterized protein	5.04E-04	-1.58
E2R974	FABP4	Adipocyte-type fatty acid-binding protein	2.55E-07	-1.53
D6BR72	KRT71	Keratin 71	-	-1.50
F1PK58	OLFML1	Olfactomedin like 1	5.80E-08	-1.49
F1PDZ7	CA2	Carbonic anhydrase	3.91E-02	-1.46
F1PDJ0	APOC3	Apolipoprotein C-III	3.61E-04	-1.45

Proteins are ranked by their log2 fold change CAS/normal stroma. Up-regulated proteins have a positive value for log2 ratio, down-regulated ones display a negative value for log2 ratio. Proteins without a p-value were exclusively detected in CAS (for the up-regulated ones) or normal stroma (for the down-regulated ones), wherefore no p-value could be calculated.

heavily counterbalanced by the disadvantages that arise through the need of a high grade of coordination to achieve proper collection and storage. Furthermore, tissue morphology of fresh-frozen tissue is inferior to that of FFPE. And finally, using fresh-frozen tissue precludes the use of most archival samples, which are generally processed as FFPE. Indeed, proteomic analysis has also been shown to work for microdissected FFPE tissue [35,36]. This possibility to analyze specific areas of archival patient samples by both RNAseq and proteomics unlocks a novel dimension of hard-to-analyze samples for investigation.

Our approach revealed several interesting insights into CAS-reprogramming. Importantly, the most significantly deregulated proteins are all produced by stromal cells, thus validating our approach to isolate and analyze stroma. Some of the proteins, such as Collagen type VIII, Fibrillin and LTBP2, are known to be of mainly stromal origin and are upregulated in

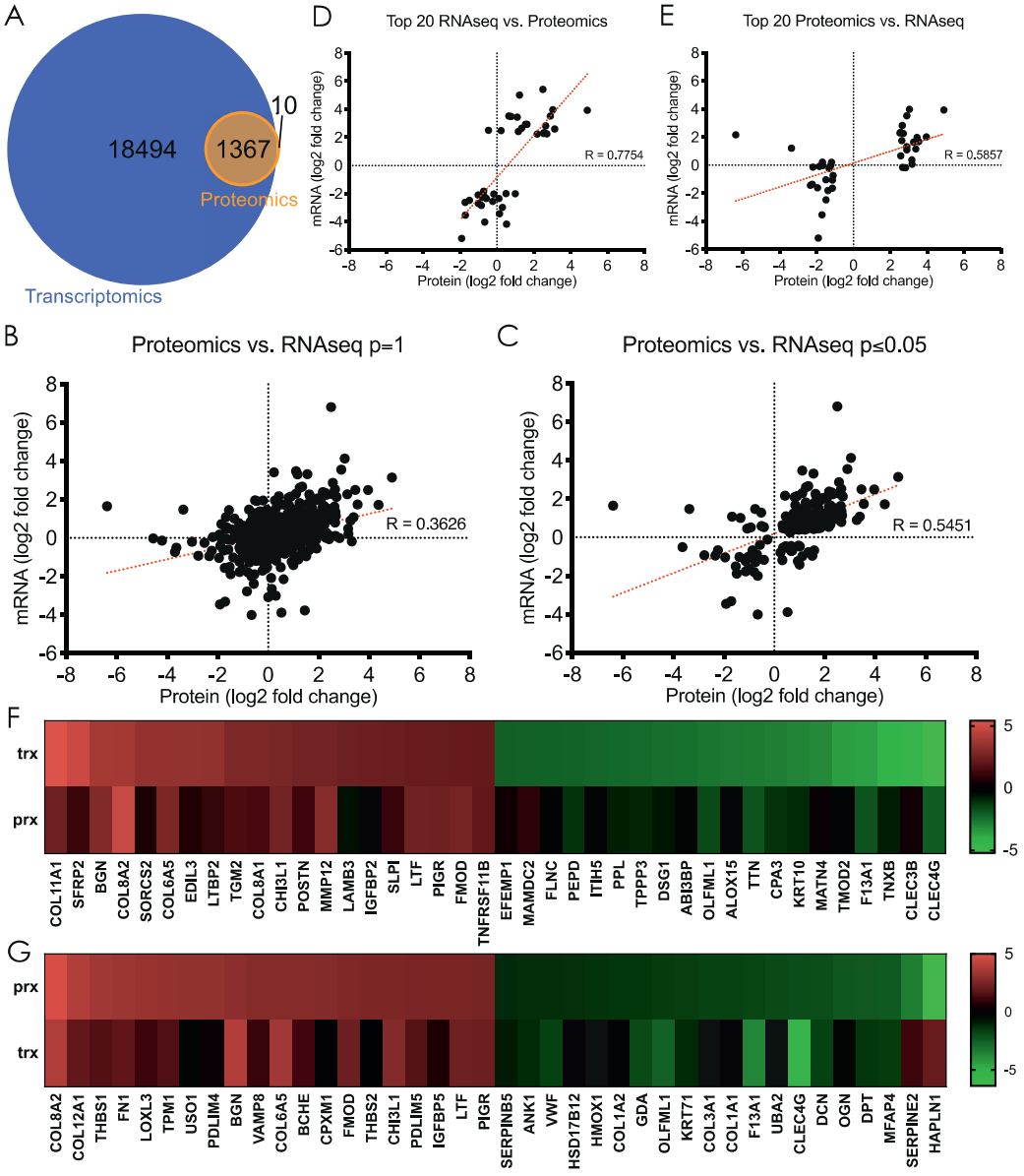


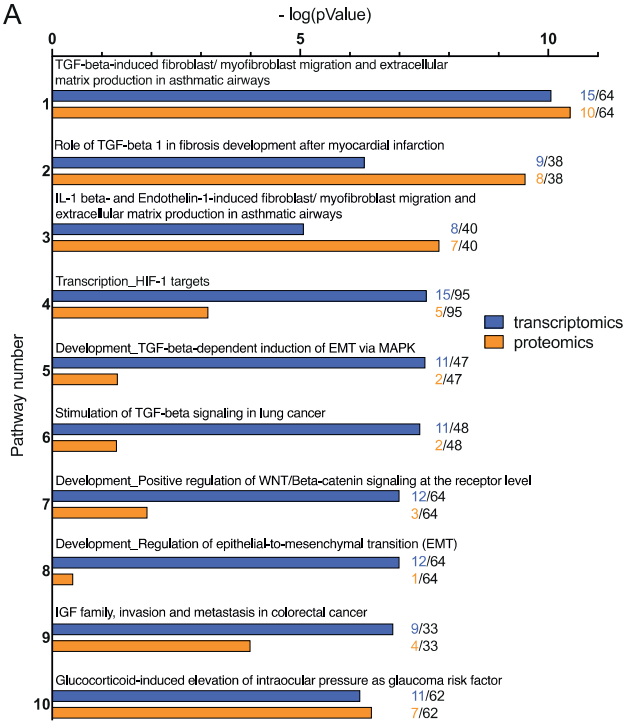
Fig. 3. Comparison of proteomic and transcriptomic profiling of stromal reprogramming. A) Venn diagram showing the overlap between the proteomic (orange) and the transcriptomic (blue) dataset. B) and C) Correlation analysis of the expression changes between CAS and normal stroma for the targets identified through both transcriptomics and proteomics for all data points (adjusted p-value threshold 1, B) and for the significantly changing data points (adjusted p-value threshold 0.05, C). D and E) Linear regression analysis of the top 20 up- and down-regulated targets from the RNAseq dataset with the respective proteomic data (D) and from the proteomic dataset with the respective RNAseq data (E). F and G) Heatmap of top 20 up- and down-regulated targets from D (F) and E (G).

colorectal cancer [37,38]. Especially the composition of collagens seemed to shift quite strongly between normal stroma and CAS. The most significantly upregulated collagens that we found in CAS were the following: COL4A1, COL6A3, COL8A2, COL11A1, COL11A2, COL12A1, and COL15A1. Most of the collagens are produced by stromal cells, and fibrillar collagens and collagen VI for instance are collagens known to be produced by CAFs (reviewed in [71]). Collagens have been found to have important roles in different types of cancer, for example collagen XII in gastric and ovarian cancer [39–41], or collagen XI in [42–45]. Additionally, genes of collagen XI and XII are also among the most upregulated genes in CAS from human breast cancer [46]. Hence, the up-regulation of these particular collagens on the protein level is in strong accordance with the findings of other studies.

The strong involvement of the tumor necrosis factor (TNF) superfamily, with the regulating molecules THBS1, LTF, ACP5, and BPI is in accordance with its known role. TNF is frequently upregulated in human epithelial malignancies such as breast cancer [47–49]. On one hand, TNF affects tumor proliferation and survival, EMT, metastasis and recurrence (reviewed in [50]). On the other hand, TNF possesses some anticancer properties through inducing cancer cell death [51]. Further analysis of the involvement of the TNF pathway in CAS of canine and human mCA will likely yield interesting data.

Comparative studies have found that correlations between mRNA and protein levels in model organisms can be relatively weak and uncertain or moderately positive [26]. In our dataset, the correlation coefficient between

A



B

Pathway no.	Proteomics	Transcriptomics
1	COL1A1, Tenascin-C, ITGB1, Collagen I, PAI1, COL4A1, Biglycan, COL1A2, Collagen IV, Decorin	ITGA1, Biglycan, TGF-beta 2, TGF-beta 3, Fibronectin, Tenascin-C, ITGB1, HAS2, PAI1, COL4A1, TIMP1, TGF-beta receptor type I, COL5A1, Collagen IV, Decorin
2	COL1A1, Tenascin-C, Collagen I, PAI1, Prolyl endopeptidase, Thrombospondin 1, Biglycan, COL1A2	Thrombospondin 1, Biglycan, Fibronectin, Tenascin-C, ACTA2, PAI1, EDNRA, TIMP1, TGF-beta receptor type I
3	COL1A1, Collagen I, PAI1, COL4A1, Thrombospondin 1, COL1A2, Decorin	Thrombospondin 1, Fibronectin, HAS2, PAI1, COL4A1, EDNRA, TIMP1, Decorin
4	PAI1, PGK1, ENO1, Heme oxygenase 1, Thrombospondin 1	ROR-alpha, TGM2, Adrenomedullin, DEC1 (Stra13), Thrombospondin 1, TGF-beta 2, TGF-beta 3, VEGFR-1, PAI1, CX3CR1, Stanniocalcin 2, CXCR4, LOXL2, IBP3, Transferrin
5	ITGB1, PAI1	E-cadherin, TGF-beta 2, TGF-beta 3, Fibronectin, ITGB1, TGF-beta, alpha-V/beta-1 integrin, Claudin-1, PAI1, TGF-beta receptor type I, c-Fos
6	ITGB1, PAI1	SCUBE3, E-cadherin, TGF-beta 2, TGF-beta 3, Fibronectin, ITGB1, ACTA2, TGF-beta, PAI1, TGF-beta receptor type I, Tropomyosin-2
7	Tenascin-C, Biglycan, Filamin A	LGR6, Biglycan, Tenascin-C, Glypican-3, DKK2, SFRP2, WNT5B, WNT, FZD2, Tenascin-N, Frizzled, SFRP1
8	PAI1	E-cadherin, TGF-beta 2, TGF-beta 3, Fibronectin, ACTA2, Claudin-1, WNT, PAI1, Frizzled, EDNRA, Sno-N, TGF-beta receptor type I
9	ITGB1, Collagen I, PAI1, Collagen IV	E-cadherin, Fibronectin, ITGB1, alpha-V/beta-6 integrin, PAI1, IGF-2, alpha-1/beta-1 integrin, ITGAV, Collagen IV
10	ITGB1, PAI1, COL4A1, Filamin B, Thrombospondin 1, Filamin A, Collagen IV	Thrombospondin 1, Fibronectin, ITGB1, Elastin, Alpha-actinin, PAI1, COL4A1, LAMA1, MLCK, LAMB3, Collagen IV

Fig. 4. Comparable picture of pathway activation between transcriptomic and proteomic data. A) MetaCore pathway analysis of the human orthologues of all proteomic and transcriptomic targets, respectively, showing the top 10 activated pathways. B) Targets assigned to the activated pathways as shown in A).

transcriptomics and proteomics was R=0.3626, similar to what has been reported for other tissues before [17–19]. Setting the threshold of the adjusted p-value at 0.05 increased the correlation coefficient to 0.5451. Similarly, restricting the analysis to the top 20 targets from either the transcriptomic or proteomic dataset further increased the correlation, suggesting that the more strongly targets are deregulated, the more consistent the correspondence between mRNA and protein levels become.

The strong role of TGF-beta related signaling that emerged from pathway analysis is well supported by literature. TGF-beta naturally possesses both tumor-suppressive as well as tumor-promoting qualities, depending on the context. As tumor-promoter TGF-beta can support the promotion of tumor growth and invasion, evasion of immune surveillance, and metastasis [52,53]. Also associated with the TGF-beta pathway is PAI-1 (plasminogen activator inhibitor-1), a serine protease inhibitor that is strongly upregulated in CAS. PAI-1 inhibits tissue plasminogen activator and urokinase, the activators of plasminogen and hence fibrinolysis. In cancer, it promotes its invasion and metastasis and correlates with poor prognosis in breast cancer [54–56]. In summary, these results suggest that, while the proteomics-based analysis of LCM-FFPE tissue yields a smaller set of detected targets, the protein and transcript levels show a substantial degree of correlation, especially for the most deregulated targets. Moreover, the results from both approaches reveal a comparable picture with regards to pathway activation during stromal reprogramming in canine mCA.

By leveraging the relevance of canine CAS as a model of the human disease, we establish increased expression of LTBP2, IGFBP2, COL6A5, POSTN, FN1, COL4A1, COL12A1, PLOD2, COL4A2, and IGFBP7 to be strongly correlated with worse overall survival for human breast cancer patients. Validation of increased Collagen IV and Fibronectin on protein

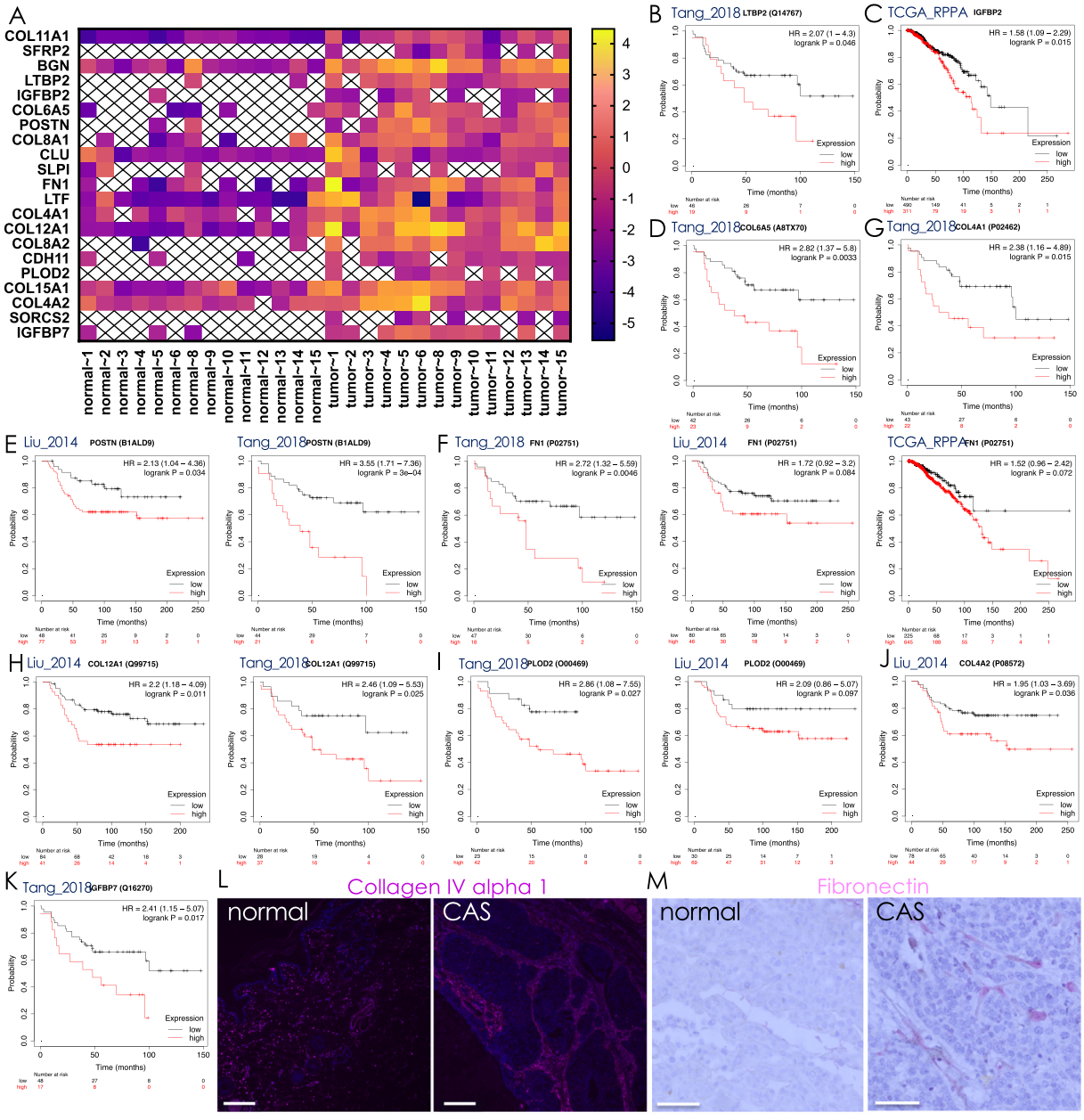


Fig. 5. Proteomic CAS dataset validates deregulation of stromal targets with prognostic value for human breast cancer. A) Heatmap of protein expression of 21 potential disease-promoting targets as identified by RNAseq. B – K) Kaplan-Meier curves of overall survival for the indicated proteins listed in A) to visualize survival differences between high and low-expressing patients (auto select best cutoff) with a follow-up of up to 150 months (dataset Tang_2018) or 250 months (datasets Liu_2014 and TCGA_RPPA). B) LTBP2, C) IGFBP2, D) COL6A5, E) POSTN, F) FN1, G) COL4A1, H) COL12A1, I) PLOD2, J) COL4A2 and K) IGFBP7. L) Immunofluorescent staining of Collagen IV (purple), in CAS and matched normal stroma of a representative canine simple mCA sample. Dapi (blue) stains cell nuclei, and marks largely blue areas in the right panels as tumor cells, while demarcating normal mammary glands in the panels on the left. Scale bar = 250 μm. M) Immunohistochemistry of fibronectin (pink) in CAS and matched normal stroma of a representative canine simple mCA sample. Stroma between normal glands is negative, while stroma next to tumor cells clearly stains positive. Scale bar = 50 μm.

level in canine CAS further substantiates these findings. Of note, PDGFRB and FAP, two CAS markers that we have previously validated as upregulated using IHC [13] were also found to be upregulated using proteomics. LTBP2 is critically involved in regulation of TGF-beta signaling [57] and its expression is significantly elevated in human breast cancer in correlation with clinical stage and other adverse prognostic factors [58]. As a regulator of PI3K signaling, the expression IGFBP2 is strongly correlated with grade of malignancy in many tumors and especially in breast cancer [59]. Collagen VI has been recently identified as driver of invasion and metastasis of breast cancer and high protein levels of Collagen VI have been found by bulk tumor proteomic analysis of more than 500 human cancers [60–62]. Strongly elevated levels of both POSTN and FN1 are well documented in a wide variety of tumors, including breast cancer, and have been shown to promote tumor invasion and metastasis through manifold effects on different cancer hallmarks (reviewed in [63,64]). Similarly, Collagen IV is positively correlated with larger and more aggressive breast tumors [65] and has been shown to

promote cancer cell invasion and migration [66]. High PLOD2 expression is associated with increased mortality risk in breast cancer [67]. Functionally, it promotes fibrillar collagen formation and thereby increases tumor stiffness and is required for metastasis. IGFBP7 expression in tumor cells has been suggested to be anti-neoplastic [68]. However, when expressed in CAS, it has growth-promoting effects on tumor cells through a paracrine signaling mechanism that can promote anchorage-independent growth of tumor cells [69]. Interestingly, IGFBP7 also binds to Collagen IV [68].

The fact that on transcriptomics level only one of the top 10 adverse genes [12] actually correlates with worse OS, while on the protein level there are 4 among these top 10 targets that predict worse OS suggests that proteomic changes could potentially be a better predictor than transcriptomic changes, at least when it comes to analyzing expression of stromal genes in bulk tissue. This might be due to the fact that RNAseq data usually turns back much more differentially expressed targets than proteomics analysis (in this study we detected 13% of all transcripts on the proteomic level). Important transcriptomic changes in single targets are therefore more likely to be 'overshadowed' by other targets that might not have anything to do with OS in contrast to the relatively 'smaller' proteomic dataset. Another possibility is, that proteomic detection could be biased towards the most abundant proteins, which might also hold more biological relevance. While it's currently not possible to finally answer this question, this clearly further underlines the point that transcriptomic and proteomic changes are not always in direct concordance and that proteomic profiling of distinct tissue compartments can provide novel insight into biological questions that go beyond validation of RNAseq data.

Collectively, the results presented in this paper will serve as a starting point for mechanistic follow-up studies to further delineate and strengthen the role of these stromal genes in development and progression of canine and human mCA. Hence by validating the deregulation of previously identified differentially expressed transcripts on protein level, our proteomic analysis of CAS reprogramming in canine mCA further substantiates these targets as disease-modulating stromal components with implications for breast cancer in both humans and dogs.

Data and code availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [70] partner repository with the dataset identifier PXD023023. Raw and processed data from RNAsequencing reported in this study are available at the Gene Expression Omnibus under accession number GSE135183. All other data supporting our findings is contained in the manuscript and in the supplementary figures and tables.

Author contributions

A.P. performed LCM and data analysis with the help of E.B. L.K. isolated proteins from the LCM samples and performed LC-MS/MS. P.A. provided technical and analytical input. A.M. is a board-certified veterinary pathologist, and performed and supervised choice of clinical cases. F.G. is nationally certified pathologist and provided input regarding selection of areas of interest by LCM. E.M. was responsible for study design, supervision, data analysis and funding. A.P., E.B. and E.M. wrote the first draft of the manuscript. All authors read, contributed to, and approved the final manuscript.

Acknowledgments

The authors thank the histology laboratory of the Institute of Veterinary Pathology, University of Zürich for slide preparation and technical assistance, as well as Lennart Opitz and Dr. Witold Wolski (Functional Genomics Center Zürich) for their expertise regarding proteomics and transcriptomic analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neo.2021.03.001.

References

1 Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. *Nat Rev Drug Discov* 2019:1–17.

2 Gandellini P, Andriani F, Merlino G, D'Aiuto F, Roz L, Callari M. Complexity in the tumour microenvironment: cancer associated fibroblast gene expression patterns identify both common and unique features of tumour-stroma crosstalk across cancer types. *Semin Cancer Biol* 2015;**35**:96–106. doi:10.1016/j.semcancer.2015.08.008.

3 Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* 2012;**21**:309–22.

4 Gkretsi V, Stylianopoulos T. Cell adhesion and matrix stiffness: coordinating cancer cell invasion and metastasis. *Front Oncol* 2018;**8**:145.

5 Liu D, Xiong H, Ellis AE, Northrup NC, Rodriguez CO, O'Regan RM, Dalton S, Zhao S. Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer. *Cancer Res* 2014;**74**:5045–56.

6 Queiroga FL, Raposo T, Carvalho MI, Prada J, Pires I. Canine mammary tumours as a model to study human breast cancer: most recent findings. *In Vivo* 2011;**25**:455–65.

7 Schiffman JD, Breen M. Comparative oncology: what dogs and other species can teach us about humans with cancer. *Philos Trans R Soc Lond B Biol Sci* 2015;**370**.

8 Salas Y, Márquez A, Diaz D, Romero L. Epidemiological study of mammary tumors in female dogs diagnosed during the period 2002-2012: a growing animal health problem. *PLoS One* 2015;**10**:e0127381.

9 Goldschmidt M, Pena L, Rasotto R, Zappulli V. Classification and grading of canine mammary tumors. *Vet Pathol* 2011;**48**:117–31.

10 Amini P, Ettlin J, Opitz L, Clementi E, Malbon A, Markkanen E. An optimised protocol for isolation of RNA from small sections of laser-capture microdissected FFPE tissue amenable for next-generation sequencing. *BMC Mol Biol* 2017;**18**:22.

11 Amini P, Nassiri S, Ettlin J, Malbon A, Markkanen E. Next-generation RNA sequencing of FFPE subsections reveals highly conserved stromal reprogramming between canine and human mammary carcinoma. *Dis Model Mech* 2019;**12**(8). doi:10.1242/dmm.040444.

12 Amini P, Nassiri S, Malbon A, Markkanen E. Differential stromal reprogramming in benign and malignant naturally occurring canine mammary tumours identifies disease- modulating stromal components. *Sci Rep* 2020:1–13.

13 Ettlin J, Clementi E, Amini P, Malbon A, Markkanen E. Analysis of gene expression signatures in cancer-associated stroma from canine mammary tumours reveals molecular homology to human breast carcinomas. *Int J Mol Sci* 2017;**18**:1101–19.

14 Markkanen E. Know thy model: charting molecular homology in stromal reprogramming between canine and human mammary tumors. *Front Cell Dev Biol* 2019;**7**:348.

15 Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, Chen H, Omeroglu G, Meterissian S, Omeroglu A, et al. Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med* 2008;**14**:518–27.

16 Pepin F, Bertos N, Laferriere J, Sadekova S, Souleimanova M, Zhao H, Finak G, Meterissian S, Hallett MT, Park M. Gene expression profiling of microdissected breast cancer microvasculature identifies distinct tumor vascular subtypes. *Breast Cancer Res* 2012;**14**:R120.

17 Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* 2003;**4**:117–18.

18 Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett* 2009;**583**:3966–73.

19 Popovic D, Koch B, Kueblbeck M, Ellenberg J, Pelkmans L. Multivariate control of transcript to protein variability in single mammalian cells. *Cell Systems* 2018;**7**:398–411 e6.

20 Braakman RBH, Stingl C, Tilanus-Linthorst MMA, van Deurzen CHM, Timmermans MAM, Smid M, Foekens JA, Luider TM, Martens JWM, Umar A. Proteomic characterization of microdissected breast tissue environment provides a protein-level overview of malignant transformation. *Proteomics* 2017;**17** 1600213–11.

21 Türker, C., Akal, F., Panse, C., joho, D., Barkow-Oesterreicher, S., Rehrauer, H., and Schlapbach, R. (2011). B-Fabric: the swiss army knife for life sciences. Proceedings of the Th International Conference on Extending Database Technology, Lausanne, Switzerland 1–5.

22 Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008;**26**:1367–72.

23 Bates, D., Mächler, M., Ben Bolker, and Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv stat.CO.

24 Kuznetsova, A., Brockhoff, P.B., and Christensen, R. lmerTest package: tests in linear mixed effects models. *J Statist. Software* December 2017;**82**(13):1–26.

25 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Statist Soc. Series B (Methodological)* 1995;**57**(1):289–300.

26 Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 2012;**13**:227–32.

27 Győrffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* 2009;**123**:725–31.

28 Akbani R, Ng PKS, Werner HMJ, Shahmoradgoli M, Zhang F, Ju Z, Liu W, Yang J-Y, Yoshihara K, Li J, et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat Commun* 2014;**5** 3887–15.

29 Liu NQ, Stingl C, Look MP, Smid M, Braakman RBH, De Marchi T, Sieuwerts AM, Span PN, Sweep FCGJ, Linderholm BK, et al. Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer. *J Natl Cancer Inst* 2014;**106** 5287–10.

30 Tang, W., Zhou, M., Dorsey, T.H., Prieto, D.A., Wang, X.W., Ruppin, E., Veenstra, T.D., and Ambs, S. Integrated proteotranscriptomics of breast cancer reveals globally increased protein- mRNA concordance associated with subtypes and survival. *Genome Medicine* 2018;**10**(94):1–14.

31 Hedegaard J, Thorsen K, Lund MK, Hein A-MK, Hamilton-Dutoit SJ, Vang S, Nordentoft I, Birkenkamp-Demtröder K, Kruhøffer M, Hager H, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS One* 2014;**9**:e98187.

32 Sinicropi D, Qu K, Collin F, Crager M, Liu M-L, Pelham RJ, Pho M, Dei Rossi A, Jeong J, Scott A, et al. Whole transcriptome RNA-Seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue. *PLoS One* 2012;**7**:e40092.

33 Tanca A, Pagnozzi D, Burrai GP, Polinas M, Uzzau S, Antuofermo E, Addis MF. Comparability of differential proteomics data generated from paired archival fresh-frozen and formalin-fixed samples by GeLC–MS/MS and spectral counting. *J Proteomics* 2012;**77**:561–76.

34 Gomig THB, Cavalli IJ, de Souza RLR, Vieira E, Lucena ACR, Batista M, Machado KC, Marchini FK, Marchi FA, Lima RS, et al. Quantitative label-free mass spectrometry using contralateral and adjacent breast tissues reveal differentially expressed proteins and their predicted impacts on pathways and cellular functions in breast cancer. *J Proteomics* 2019;**199**:1–14.

35 Heaton KJ, Master SR. Peptide extraction from formalin-fixed paraffin-embedded tissue. *Curr Protoc Protein Sci* 2011;**1**:5–23 Unit235.19.

36 Smith AL, Sun M, Bhargava R, Stewart NA, Flint MS, Bigbee WL, Krivak TC, Strange MA, Cooper KL, Zorn KK. Proteomic analysis of matched formalin-fixed, paraffin-embedded specimens in patients with advanced serous ovarian carcinoma. *Proteomes* 2013;**1**:240–53.

37 Drev D, Bileck A, Erdem ZN, Mohr T, Timelthaler G, Beer A, Gerner C, Marian B. Proteomic profiling identifies markers for inflammation-related tumor–fibroblast interaction. *Clin Proteomics* 2017:1–16.

38 Puré E, Blomberg R. Pro-tumorigenic roles of fibroblast activation protein in cancer: back to the basics. *Oncogene* 2018:1–15.

39 Duan S, Gong B, Wang P, Huang H, Luo L, Liu F. Novel prognostic biomarkers of gastric cancer based on gene expression microarray: COL12A1, GSTA3, FGA and FGG. *Mol Med Rep* 2018;**18**:3727–36.

40 Januchowski R, Świerczewska M, Sterzyńska K, Wojtowicz K, Nowicki M, Zabel M. Increased expression of several collagen genes is associated with drug resistance in ovarian cancer cell lines. *J Cancer* 2016;**7**:1295–310.

41 Jiang X, Wu M, Xu X, Zhang L, Huang Y, Xu Z, He K, Wang H, Wang H, Teng L. COL12A1, a novel potential prognostic factor and therapeutic target in gastric cancer. *Mol Med Rep* 2019;**20**:3103–12.

42 Freire J, Domínguez-Hormaetxe S, Pereda S, De Juan A, Vega A, Simón L, Gómez-Román J. Collagen, type XI, alpha 1: An accurate marker for differential diagnosis of breast carcinoma invasiveness in core needle biopsies. *Pathology Res Pract* 2014;**210**:879–84.

43 Halsted KC, Bowen KB, Bond L, Luman SE, Jorcyk CL, Fyffe WE, Kronz JD, Oxford JT. Collagen α1(XI) in normal and malignant breast tissue. *Modern Pathol* 2008;**21**:1246–54.

44 Shen L, Yang M, Lin Q, Zhang Z, Zhu B, Miao C. COL11A1 is overexpressed in recurrent non-small cell lung cancer and promotes cell proliferation, migration, invasion and drug resistance. *Oncol Rep* 2016;**36**:877–85.

45 Wu Y-H, Chang T-H, Huang Y-F, Huang H-D, Chou C-Y. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene* 2014;**33**:3432–40.

46 Ma X-J, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res* 2009;**11**:46.

47 Boldrini, L.E.A. (2017). Tumour necrosis factor-a and transforming growth factor-b are significantly associated with better prognosis in non-small cell lung carcinoma: putative relation with BCL-2-mediated neovascularization. *British J Cancer.* 2000;**83**(4):1–7, 480486.

48 Katerinaki E, Evans GS, Lorigan PC, MacNeil S. TNF-α increases human melanoma cell invasion and migration in vitro: The role of proteolytic enzymes. *Br J Cancer* 2003;**89**:1123–9.

49 Miles DW, Happerfield LC, Naylor MS, Bobrow LG, Rubens RD, Balkwwill FR. Expression of tumour necrosis factor (TNFα) and its receptors in benign and malignant breast tissue. *Int J Cancer* 1994;**56**:777–82.

50 Cruceriu, D., Baldasici, O., Balacescu, O., and Berindan-Neagoe, I. The dual role of tumor necrosis factor-alpha (TNF-α) in breast cancer: molecular insights and therapeutic approaches. *Cellular Oncol* 2020;**43**(1):1–18.

51 Wang X, Lin Y. Tumor necrosis factor and cancer, buddies or foes? *Acta Pharmacol Sinica* 2008;**29**:1275–88.

52 Korpal M, Kang Y. Targeting the transforming growth factor-β signalling pathway in metastatic cancer. *Eur J Cancer* 2010;**46**:1232–40.

53 Massagué J. TGFβ in cancer. *Cell* 2008;**134**:215–30.

54 Duffy MJ. The urokinase plasminogen activator system: role in malignancy. *Curr Pharm Des* 2004;**10**:39–49.

55 Duffy MJ, McGowan PM, Harbeck N, Thomssen C, Schmitt M. uPA and PAI-1 as biomarkers in breast cancer: validated for clinical use in level-of-evidence-1 studies. *Breast Cancer Res* 2014;**16** 428–10.

56 Wei X, Li S, He J, Du H, Liu Y, Yu W, Hu H, Han L, Wang C, Li H, et al. Tumor-secreted PAI-1 promotes breast cancer metastasis via the induction of adipocyte-derived collagen remodeling. *Cell Commun Signal* 2019;**17** 58–18.

57 Robertson IB, Horiguchi M, Zilberberg L, Dabovic B, Hadjiolova K, Rifkin DB. Latent TGF-β-binding proteins. *Matrix Biol* 2015;**47**:44–53.

58 Gu CJ, Jin Q, Liu G, Ni K, Ni QC. [The expression of LTBP2 in breast cancer and its clinical significance]. *Zhonghua Yi Xue Za Zhi* 2018;**98**:264–8.

59 Hoeflich A, Russo VC. Physiology and pathophysiology of IGFBP-1 and IGFBP-2 - Consensus and dissent on metabolic control and malignant potential. *Best Pract Res Clin Endocrinol Metab* 2015;**29**:685–700.

60 Cescon M, Gattazzo F, Chen P, Bonaldo P. Collagen VI at a glance. *J Cell Sci* 2015;**128**:3525–31.

61 Chen F, Chandrashekar DS, Varambally S, Creighton CJ. Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. *Nat Commun* 2019;**10** 5679–15.

62 Wishart AL, Conner SJ, Guarin JR, Fatherree JP, Peng Y, McGinn RA, Crews R, Naber SP, Hunter M, Greenberg AS, et al. Decellularized extracellular matrix scaffolds identify full-length collagen VI as a driver of breast cancer cell invasion in obesity and metastasis. *Sci Adv* 2020;**6**:eabc3175.

63 Efthymiou G, Saint A, Ruff M, Rekad Z, Ciais D, Van Obberghen-Schilling E. Shaping up the tumor microenvironment with cellular fibronectin. *Front Oncol* 2020;**10**:641.

64 González-González L, Alonso J. Periostin: a matricellular protein with multiple functions in cancer development and progression. *Front Oncol* 2018;**8**:225.

65 Ioachim E, Charchanti A, Briasoulis E, Karavasilis V, Tsanou H, Arvanitis DL, Agnantis NJ, Pavlidis N. Immunohistochemical expression of extracellular matrix components tenascin, fibronectin, collagen type IV and laminin in breast cancer: their prognostic value and role in tumour invasion and progression. *Eur J Cancer* 2002;**38**:2362–70.

66 Xu S, Xu H, Wang W, Li S, Li H, Li T, Zhang W, Yu X, Liu L. The role of collagen in cancer: from bench to bedside. *J Transl Med* 2019;**17**:309–22.

67 Gilkes DM, Bajpai S, Wong CC, Chaturvedi P, Hubbi ME, Wirtz D, Semenza GL. Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. *Mol Cancer Res* 2013;**11**:456–66.

68 Jin L, Shen F, Weinfeld M, Sergi C. Insulin Growth Factor Binding Protein 7 (IGFBP7)-Related Cancer and IGFBP3 and IGFBP7 Crosstalk. *Front Oncol* 2020;**10**:727.

69 Rupp C, Scherzer M, Rudisch A, Unger C, Haslinger C, Schweifer N, Artaker M, Nivarthi H, Moriggl R, Hengstschläger M, et al. IGFBP7, a novel tumor stroma marker, with growth-promoting effects in colon cancer through a paracrine tumor-stroma interaction. *Oncogene* 2015;**34**:815–25.

70 Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 2019;8;47(D1):D442–D450.

71 Nissen, N.I., Karsdal, M., and Willumsen, N. Collagens and cancer associated fibroblasts in the reactive stroma and its relation to cancer biology. *J Exper Clinical Cancer Res*, 2019;**38**:115, 1–12.

72 Yates AD, et al. 'Ensembl 2020'. *Nucleic Acids Res* 2020;**48**(D1):D682–8. doi:10.1093/nar/gkz966.